

# Aprendizagem Automática

Semestre de Inverno 2023/2024

G. Marques

# Aprendizagem Automática

## O que significa?

- Algoritmos que, para uma dada tarefa, “aprendem” com os dados e melhoram com experiência sem necessitarem de ser explicitamente programados.
- ... o que geralmente implica definir um modelo matemático, uma medida de desempenho, e usar os dados para adaptar os parâmetros do modelo de modo a melhorar o desempenho.

# Aprendizagem Automática

## Tipos de Aprendizagem

- **Supervisionada**

Fornece-se ao algoritmo exemplos e as respetivas saídas desejadas. O objetivo é estimar o mapeamento entre os exemplos de entrada e as saídas desejadas.

- ▶ **Classificação:** Os dados estão divididos em duas ou mais classes e o objetivo é decidir a quais classes novos dados pertencem.
- ▶ **Regressão:** Prever o valor de uma dada variável baseado nos valores de outras.

- **Não Supervisionada**

Os dados não estão etiquetados: não há classes nem saídas desejadas. O objetivo é encontrar padrões (estruturas, repetições, ordem) nos dados. Duas das principais aplicações são *clustering* (agrupamentos) e redução da dimensão dos dados.

# Aprendizagem Automática

## Tipos de Aprendizagem

- Semi Supervisionada

Mistura das duas anteriores.

- Reforço

Agentes que interagem com um ambiente dinâmico de modo às ações tomadas produzirem uma “recompensa” (jogos, robótica).

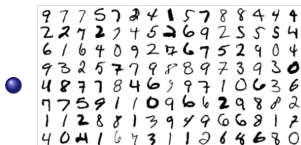
# Aprendizagem Automática

## Problemas abordados

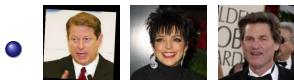
A disciplina é **uma disciplina prática** na qual são abordados diversos problemas reais. A análise de problemas, a modelação e simulação são feitas recorrendo à linguagem de programação **Python**.



Iris dataset



MNIST dataset



LFW dataset

- Críticas de cinema (imDB)
- Breast Cancer Wisconsin Diagnostic dataset
- Boston House Prices dataset
- Segmentos de áudio de notas musicais. (Elect. Music Studios - U. Iowa)
- ...

# Aprendizagem Automática

## Recomendações

Para um bom aproveitamento à disciplina recomenda-se:

- Bases sólidas em álgebra linear.
- Bases sólidas em probabilidades e estatística.

Adicionalmente é necessário competências em programação em particular na linguagem **Python** (versão  $\geq 3$ ).

- Numpy + Scipy
- matplotlib
- scikit-learn
- nltk
- opencv
- ...

Instalar igualmente o [Jupyter Notebook](#)

# Aprendizagem Automática

## Programa da disciplina

### 1 Apresentação

Revisões: geração e transformação de dados multivariados.

### 2 Sistemas de Classificação

- ▶ Classificação multi-classe, multi-label
- ▶ Métricas de avaliação
- ▶ Metodologias de teste
- ▶ Métricas de distância aplicadas à classificação

Classificações baseadas em distâncias.

Classificadores de distância ao centroide e dos  $k$ -vizinhos mais próximos.

### 3 Pré-Processamento e Transformações Lineares

- ▶ Pré-processamento e normalização de dados
- ▶ Análise em componente principais (PCA)
- ▶ Discriminantes de Fisher (LDA)

Implementações de PCA e LDA em `scikit-learn`.

Aplicação: pré-processar e classificar imagens de dígitos manuscritos.

# Aprendizagem Automática

## Programa da disciplina (continuação)

### 4 Modelos Lineares

- ▶ Regressão linear e polinomial
- ▶ Discriminantes lineares e polinomiais
- ▶ Sobre aprendizagem e técnicas de regularização

Regressão e classificação com `scikit-learn` e técnicas de regularização *ridge* e *lasso*. Análise do Boston dataset e the sinais audio de instrumentos musicais.

### 5 Discriminantes Logísticos

- ▶ Descida de gradiente
- ▶ Adaptação dos parâmetros do discriminante

Generalizações de discriminantes lineares e outros classificadores em `scikit-learn`.

### 6 Trabalhar com dados de Texto

- ▶ Contagens e modelo tf-idf
- ▶ n-gramas
- ▶ Questões práticas

Analisar based de dados imdb com 50,000 críticas de filmes.

### 7 Métodos de Agrupamento

- ▶ Algoritmos k-médias e dendrogramas



# Aprendizagem Automática

## Avaliação dos Conhecimentos

A avaliação à disciplina de Aprendizagem Automática abrange três componentes:

- Fichas de exercícios para fazer em casa (individual).
- Fichas (programas em Python) na aula (individual)
- Projeto final (nota  $\geq 9.5$  valores).  
Grupos de um ou de dois alunos.  
Avaliado por discussão oral sobre o projeto e matéria da disciplina.

A nota final é obtida segundo:

$$NF = 0.50 \times (\text{Projeto} + \text{Discussão}) + 0.25 \times \text{Programas} + 0.25 \times \text{Fichas}$$

Para obter aprovação:  $NF \geq 9.5$ .

# Aprendizagem Automática

## Bibliografia

- A. Müller, S. Guido, A introduction to machine learning with Python, O'Reilly, 2017.
- [Scikit Learn Users Guide](#).
- R. Duda, P. Hart and D. Stork, Pattern Classification, John Wiley and sons, 2001.
- C. Bishop, Pattern Recognition and Machine Learning, Springer 2006.
- [D. Mackay, Information Theory, Inference, and Learning Algorithms](#), 2003.